

Is predatory publishing a real threat? Evidence from a large database study

Marcelo S. Perlin¹  · Takeyoshi Imasato¹ · Denis Borenstein¹

Received: 3 November 2017 / Published online: 19 April 2018
© Akadémiai Kiadó, Budapest, Hungary 2018

Abstract Using a database of potential, possible, or probable predatory scholarly open-access journals, the objective of this research is to study the penetration of predatory publications in the Brazilian academic system and the profile of authors in a cross-section empirical study. Based on a massive amount of publications from Brazilian researchers of all disciplines during the 2000–2015 period, we were able to analyze the extent of predatory publications using an econometric modeling. Descriptive statistics indicate that predatory publications represent a small overall proportion, but grew exponentially in the last 5 years. Departing from prior studies, our analysis shows that experienced researchers with a high number of non-indexed publications and PhD obtained locally are more likely to publish in predatory journals. Further analysis shows that once a journal regarded as predatory is listed in the local ranking system, the Qualis, it starts to receive more publications than non-predatory ones.

Keywords Predatory journals · Scholarly publishing · Open access · Lattes platform

Mathematics Subject Classification 00-01 · 99-00

✉ Marcelo S. Perlin
marcelo.perlin@ufrgs.br
Takeyoshi Imasato
t.imasato@ufrgs.br
Denis Borenstein
denis.borenstein@ufrgs.br

¹ Universidade Federal do Rio Grande do Sul, Escola de Administração, Washington Luis 855, Porto Alegre, Brazil

Introduction

The emergence of new information technologies (such as the World Wide Web) and the adoption of international scientific publication as criteria for investments, appointment and promotion by universities in many countries introduced transformations in the publishing of scholarly journals. In order to cope with these transformations, a new publishing model based on open access (OA) journals emerged (Shen and Björk 2015).

In general, OA journals are only published in electronic versions, allowing the public download of articles. In this model, the costs of publication shifts from the readers to the authors or funding agencies. The relevance of this trend can be illustrated by the increase of journals indexed in the Directory of Open Access Journals (DOAJ) (launched in 2003 at Lund University, Sweden), from 300 initial journals to the current 9900 indexed OA periodicals. This increase is consequence of the main advantage of OA journals of eliminating the price barrier for readers, improving the access to publications. The OA model has initially raised some contrary reactions from established publishers. However, the dissemination advantages of the OA publishing model has attracted the attention of several authors and the general public. More recently, some traditional publishers, such as Elsevier and Springer, have incorporated this model, offering complete open access journals or turning some prestigious journals into hybrid open access ones, where only part of the published papers are open access. Despite the heated debate over OA publishing (Abbasi 2012), Laakso (2011) point out that this model has reached a consolidated phase, and it is here to stay.

However, the incoming of OA journals had an undesirable side effect. As pointed out by Al-Khatib (2016), “some journals have been created by greedy or amateurish, entrepreneurial non-expert entities, with the sole aim of collecting article processing charges (APC) and profiting from the pockets of desperate authors”. Jeffrey Beall, an American librarian, coined the term predatory open access for publications that prioritize financial resources rather than scientific contribution (Beall 2012), and became widely known for maintaining a blog site with a list of probable predatory scholarly open-access journals from 2010 until January 2017. Although there is a controversy in the literature about the characteristics of predatory publications (Al-Khatib 2016; Nwagwu 2016), there is a consensus that these journals are ready to publish all submitted papers, without (or with questionable) peer-reviewing, as long as the authors are ready to pay the APCs (Butler 2013). Moreover, predatory journals are associated with misleading information on the Internet, including fake addresses and editorial boards (Berger and Cirasella 2015; Sorokowski et al. 2017; Kulczycki 2017). Although a small percentage of the total number of OA journals, the misconduct of predatory journals helps to put the reputation of the OA model and its practices at stake (Beall 2015b; Shen and Björk 2015), which has been the subject of many popular journals and media, such as Nature, Science, and the New York Times (Xia 2015).

The issue of predatory publishing has been discussed mainly in editorials and opinion based papers from experienced researchers. The majority of research-based papers concerning predatory publishing are based on either empirical-qualitative research, focusing on case studies, or on quantitative-based research of specific disciplines or selected outlets, measuring discrete statistics based on small databases. Therefore, there is a gap for studies simultaneously considering a massive database—including different disciplines, number of journals, and authors—and using a statistical process for estimating the relationships among variables that may influence predatory publishing. In summary, there is a lack of quantitative research towards better characterizing predatory practices (Wallace and Perri 2018).

The objective of this research is to study the penetration of predatory publications in the Brazilian academic system and the profile of authors in a cross-section empirical study. Despite a significant growth in the last decade,¹ research practices still lag behind international standards in some fields. A higher pressure for internationalization from governmental agencies allied to potential failures in quality control and evaluation make Brazilian researchers a potential target for editorial misconduct (van Noorden 2013) and predatory publishers (Shen and Björk 2015). A database of ISSNs and names of journals in Beall's list is built using a web scrapping algorithm. Predatory publications are identified using three methods based on Beall's list, the DOAJ (directory of open access), and standard measures of impact factors (SJR and JCR). The data was analyzed using an econometric model, establishing the relationship between the profile of researchers and predatory publishing. We also assess the penetration of predatory journals in Brazil's main quality assessment tool, Qualis/CAPES.

The paper is organized as follows. “Materials and methods” section discusses the origins and extraction methods of the raw dataset, followed by a description of the statistical methodology. “Results” section presents and analyzes the results. Next, we discuss the implications on the policymaking of the obtained results in “Discussion and policy implications” section. Finally, “Conclusions” section summarizes our conclusions and discusses some future research.

Literature review

As an issue that has been rising in the last decade, the literature on predatory publishing is still scant in terms of research-based publications, which is predominantly discussed in editorials, opinion papers, or conferences. As predatory publishing attracted attention as a scientific subject, attempts to understand and research this phenomena started to take place. Bowman (2014) identifies the *modus operandis* of predatory journals, alerting researchers to common practices undertaken by these journals. The practices were ratified by Djuric (2015), performing an experiment of submitting a hoax article to a possible Romanian-based predatory journal. Several characteristics raised by Beall on his blog site are also evaluated in this study. Lukić et al. (2014) carry out a survey about predatory publishing, defining a blacklist of potential, possible, or probable predatory journals. Clark and Thompson (2016) speculate the reasons why authors decide to publish in predatory journals.

The subsequent literature shifts towards experimental studies. Some research papers focus on the perception and opinion of researchers concerning predatory publishing. Omobowale et al. (2014) use a descriptive and empirical research method to analyze the popularity of predatory publications in Nigeria. Thirty in-depth interviews were conducted with Nigerian academics. The major conclusion of the study is that many local journals, based on practices such as self-publishing and poor review standards through connections and networks, were replaced by sub-standard predatory journals with international nomenclature and addresses, keeping the Nigerian academia continually peripheral. Shuva and Taisir (2016) present a similar qualitative study with academics from Bangladesh, using an online survey to collect data about the perception of faculty members on OA and predatory journals. There was a clear trend that researchers favor publishing in OA journals with the same quality and prestige of traditional ones. However, a considerable number of

¹ See AJE research report for Brazilhere.

respondents were unaware of the existence of predatory publishers altogether. Thus, there is a proposition in the literature that predatory publishing could be displacing local traditional journals in emerging economies.

Shen and Björk (2015) perform a longitudinal study of articles published between 2010 and 2014 in 613 journals from Beall's list in order to estimate discrete statistics related to the predatory publishing, such as the number of journals and articles, distribution over scientific disciplines, average number of articles per journal, country of publishers, country of authors, APC levels, and publishing speed. Concerning APC levels, Xia (2015) finds an average cost of US\$ 94 per publication, examining 297 journals listed on Beall's list. Manca et al. (2017b, 2017a) compiles a list of predatory journals on rehabilitation, neurosciences, and neurology, respectively, offering almost the same statistics by Shen and Björk (2015) for these three disciplines.

Regarding the profile of those who publishes in allegedly predatory publisher, there is a presumption in the literature that young and inexperienced academics are the major authors of this kind of outlet (Al-Khatib 2016; Frandsen 2017; Xia et al. 2015). Al-Khatib (2016) proposes actions to protect inexperienced authors against predatory journals. Frandsen (2017) studies the characteristics of authors who cite articles published in a selected potentially predatory outlets, showing that they tend to share the same profile of the publishing authors, particularly for being inexperienced authors. Additional results show that experienced authors, to a lesser extent, also cite papers published in those publications.

Xia et al. (2015) conducts a statistically-based study on authorship in predatory journals towards identifying the academic profile, based on geographical location and different groups of OA journals defined by the authors, using data from seven pharmacological journals. The main finding of this study pertains to the dominance of researchers from developing countries publishing in such journals. Accordingly, scientists in developing countries started to be regarded as the most susceptible to predatory publishing due to the globalization of the “publish and perish” motto (Omobowale et al. 2014; Nwagwu 2016). There is evidence that a considerable amount of papers published in identified predatory journals are mainly from countries in Africa (Nigeria and South Africa) and Asia (China, India, Iran, Malaysia, and Pakistan) (Omobowale et al. 2014; Xia et al. 2015; Nwagwu 2016), as the international publication experience is less established in these countries. Although some predatory publishers are based in the UK or in the US, most publishers are actually located elsewhere (Xia 2015). Pressure to publish abroad from governmental agencies, the lack of tradition in scientific publication by researchers, senior administrators, and policymakers and a more permissible attitude towards this kind of publication could be the reasons for predatory publishing to be predominant in developing countries (Shuva and Taisir 2016).

Nevertheless, publishing in allegedly predatory journals is also an issue in developed countries (Pyne 2017; Shen and Björk 2015; Wallace and Perri 2018). Pyne (2017) studies the reward of predatory publishing, using data from a small business school in Canada. Comparing the rewards of predatory and non-predatory publications on faculty promotion and allocation of research grants, the study demonstrates that the former produces greater financial rewards. Wallace and Perri (2018) examine the penetration of predatory publishing in the field of Economics. The authors identified that several journals in Beall's list are indexed in the Research Papers in Economics (RePEc) archives. However, the most interesting conclusions of this study are: (1) to observe that economists ranked in the top 5% on RePEc published in predatory journals in 2015; and (2) 50% of authors publishing in predatory journals in 2015 were from institutions located in US, Turkey, India, UK, Pakistan, Nigeria, Italy, and Malaysia.

The main idea of this paper is to go beyond editorials on scientific or no-scientific journals, based on the perspective of some experienced researchers or journalists, or research studies limited to a specific subject, to introduce a quantitative approach to the theme, based on a large database of published papers and constructed on reliable and replicable statistical-based methods. We hope our approach can raise relevant issues that can help research agencies and institutes, universities, and publishers, among others, to find proper policies to bring down predatory journals.

Materials and methods

Identifying predatory publications

One of the most difficult problems in carrying out research about predatory publishing is to find a consensual list of predatory journals. Although there is a consensus about their practices, we could not find a 100% error free blacklist. The most (in)famous and widely used by several researchers (Lukić et al. 2014; Xia 2015) was the previously-live internet address <https://scholarlyoa.com/>, which compiled a list of “potential, possible, or probable predatory scholarly open-access journals” authored by librarian Jeffrey Beall. The journals in this blacklist were characterized by some common practices, such as charging high fees, small or inexistent peer reviews, along with many others. Beall (2013, 2015a) present a complete description of what constitutes “predatory practices”.

The list was heavily criticized by the lack of scientific criteria and transparency, leading to some well-documented mistakes (Al-Khatib 2016; Strielkowski 2017). Besides, Beall is a self-acknowledged enemy of the OA movement. Beall deleted his site in January 2017, explaining his reasons in Beall (2017), being the last iteration archived in <https://bealllist.weebly.com/>. Strielkowski (2017) comments about the vacuum created by the deletion of Beall’s site. Towards filling this gap, Cabell’s International has launched its predatory journal blacklist as commercial product, containing around 4000 journals. However as pointed out by Strielkowski (2017), “if designed and presented to the academic community in its current form, Cabell’s Blacklist might not be, at the moment, a very good replacement of Beall’s List”.

Given the difficulty of defining a reliable list of predatory journals, datasets for this study are obtained from several sources. First, we gather the list of allegedly predatory journals and publishers from Beall’s list (BL). While not free of critics, BL provides a free compiled list of publishers with predatory practices widely used in previous studies (Berger and Cirasella 2015; Shen and Björk 2015; Wallace and Perri 2018), without a proper alternative to replace it (Strielkowski 2017). BL changed significantly over time. A query to web archive text shows the first version of the blog in 2012, with 8 individual journals and 43 publishers. As a comparison, the most current list, January 2017, has 1178 publishers and 1317 individual journals. In our research, we only use the latest available information from the now defunct blog.

However, using the list in a bibliometric analysis is difficult due to the lack of structured information and lack of scientific rigor. The list offers only the names and web links of allegedly predatory publishers and journals. Collecting information from the publisher or journal website demands a significant amount of manual work, which justifies the use of sampling methods found in Shen and Björk (2015). Therefore, additional efforts to complement Beall’s list were adopted. An important source of information about

reputable open access outlets is the *Directory of Open Access Journals* (DOAJ). According to its website:²

The Directory of Open Access Journals is a service that indexes high quality, peer-reviewed Open Access research journals, periodicals, and their articles' metadata. The Directory aims to be comprehensive and cover all open access academic journals that use an appropriate quality control system, and it is not limited to particular languages or subject areas. The Directory aims to increase the visibility and ease-of-use of open access academic journals regardless of their size and country of origin thereby promoting their visibility, usage, and impact.

Moreover, standard measures of a journals impact, SJR and JCR are also a good indication of a journal's reputation. Although some journals identified in Beall's list have received a JCR for some period Djuric (2015), most of the predatory ones are unlikely to have a significant audience or be cited in the long-run. Therefore, standard impact factors can help identify predatory journals within a large-scale study such as this one. We gather data about all journals listed in SJR (2015) and JCR (2014) from their websites,³ including their measure of impact and ISSNs.

As a robustness test, we implement three classification schemes of predatory journals. The first includes all journals found in Beall's list, matched by ISSN or name. The second includes all journals in the first classification that are not in DOAJ. The third and last group is set by looking for journals in the second group that are not included in SJR and JCR. The following nomenclature is used to characterize the defined groups:

BEALL	Journals that are in Beall's list, matched by ISSN or name
BEALL + DOAJ	Journals that are in group BEALL and not in DOAJ, matched by ISSN
BEALL + DOAJ + IF	Journals that are in group BEALL + DOAJ, but not in SJR or JCR, matched by ISSN

In our study, we only need the ISSN (International Standard Serial Number) and names of journals from Beall's list, which are later used to identify predatory publications. The specific format of an ISSN code⁴ allows the use of textual pattern matching tools. We use a proprietary software that looks for the textual pattern of ISSN on the first page and all internal subsequent subpages of an internet link. The software is able to effectively gather all valid ISSN after loading the target internet address. By using the written software for all internet addresses of individual journals and publishers from Beall's site, we are able to gather a wealth of information.

The computational method of looking for ISSN patterns in publisher's site, however, is not foolproof. If the journal does not have an ISSN, or if the symbol is not explicitly set in the webpage code, the software will not capture it. Also, cases of misinformed ISSNs are possible. Given these problems, we perform a manual check of the ISSN of the 100 journals with the highest number of matches in the publication data. Cases with incorrect ISSN were inserted in a whitelist. Despite its shortcomings, we believe that the benefits of using a fast, cheap and reproducible method of a web-scraping algorithm outweighs its limitations.

² <https://doaj.org/>.

³ See <http://scimagojr.com> and <http://jcr.incites.thomsonreuters.com>.

⁴ An ISSN is structured as XXXX-XXXY, where X is any number and Y is a number or a text.

The use of developed software resulted in 1136 unique ISSNs from 892 publishers, and 244 unique ISSN from individual journals. The data is publicly available as a CRAN distributed R package called *predatory* (Perlin 2017). The compiled list is available as a *.csv* file in Github.⁵ We hope that the data gathered from our algorithm helps other researchers and institutions to identify predatory practices.

Sample collection

The second source of information is a large database of publications from Lattes platform,⁶ a repository of academic curriculum managed by CNPq, the Brazilian Council of Research. The Lattes platform is widely popular in academic circles, and its information is systematically used in research assessments such as job interviews, the evaluation of research projects, and funding. In Lattes, it is possible to find updates and reliable information about the profile and research output from practically all academics in Brazil.

Using the same software described in Perlin et al. (2017), we collect data for all researchers with a PhD in all areas of science. As for publications data, the usual information was collected, such as title, name of journal, ISSN, and year. We filter out all publications before 2000 and after 2015, as well as any publication without a valid ISSN code.

Many researchers in the sample are not active or have published marginally. The presence of inactive researchers can bias our results in a negative way. Our study focuses on researchers that are constantly producing science and not those that obtained a PhD, but moved on to other career paths. In order to select only active scientists, we remove from the sample all researchers that have less than five publications with valid ISSNs, or who have not published at least one article after 2010. We find later that our results are robust to this methodological choice. In total, after filtering and cleaning the raw data, we gather 2,349,719 publications from 102,969 researchers in all areas of science.

Econometric modeling

The first objective of the study is to analyze the profile of researchers who publish in predatory journals. The information from this analysis can provide useful insights regarding the characteristics of researchers that are publishing their work—unwillingly or not—in journals with questionable practices.

In order to better understand the relationship between researcher profile and the probability of finding predatory publications, we estimated a generalized linear model (GLM) with fixed effects for the field of science,⁷ using each method of identifying predatory publications discussed in the previous section. This modeling approach was selected due to the properties of the dataset. The model is estimated with the standard maximum likelihood method and is formally defined as follows:

$$E(Dpred_i) = f(X_i) \quad (1)$$

⁵ <https://github.com/msperlin/predatory/tree/master/inst/extdata>.

⁶ <http://lattes.cnpq.br/>.

⁷ The model is estimated using R package **glmmML** (Brostrm 2017).

$$X_i = \alpha_j + \beta_1 DPhdAbroad_i + \beta_2 DGender_i + \quad (2)$$

$$\beta_3 DiffYear_i + \beta_4 NPub_i + \beta_5 NPub_WoS_i \quad (3)$$

where $Dpred_i$ is a Dummy variable that takes value one if publication i is predatory based on the considered group, and zero otherwise. $DPhdAbroad_i$ is a dummy variable that takes value one if author of publication i obtained his PhD outside of Brazil, and zero otherwise, $DGender_i$ is a dummy variable that takes value one if author of article i is male, and zero if female. Variables $DiffYear_i$, $NPub_i$ and $NPub_WoS$ measures the experience and international reach of the author. The first is the difference between year of publication of article i and the year of PhD completion, the second is the number of total publications by the author in the year of publication of article i and the last is the number of publications indexed in Web of Science (WoS) by the year of the publication.

Results from Perlin et al. (2017) suggest that researchers with a PhD obtained abroad are more likely to focus their publishing efforts in quality journals. Thus, we expect a negative coefficient for $DPhdAbroad_i$, meaning that this group of researchers are less likely to publish in predatory journals.

As for the effect of Gender, previous studies pointed out that male researchers are likely to publish a higher number of articles than female ones (Abramo et al. 2015; Brooks et al. 2014; Perlin et al. 2017; Rørstad and Aksnes 2015). This robust result from the literature suggests that the search for a higher volume of publications could lead male scholars to search for journals that have predatory practices.

The use of $YearPhD_i$, $NPub_i$ and $NPub_WoS$ are justified by an investigation on the relationship between researchers' maturity and predatory publishing. One piece of reasoning regarding the effect of career time is that fresh graduates are more likely to seek predatory publications, knowingly or not (Frandsen 2017). Higher competition in the academic job market can make young scholars seek international publications and possibly consider more obscure journals such as those pertinent to this study. In this case, we expect the coefficient for $DiffYears_i$ to be negative. However, empirical evidence from Pyne (2017) and Wallace and Perri (2018) suggests otherwise; some researchers with publications in predatory journals are not so inexperienced. Part of the analysis will be checking whether such a result remains true in our large-scale study.

Our study also looks at the extent of predatory publication in Qualis, the local ranking system for scientific publications. Qualis is managed by CAPES, another government agency attached to the Ministry of Education. Qualis is built by a group of selected researchers within each area, and is released on a yearly basis. It provides a ranking system wherein journals are classified in one of the following categories: A1, A2, B1, B2, B3, B4, B5 and C. The classification differs for different disciplines. As an example, using the latest Qualis (2013-2016), journal FOOD CHEMISTRY (0308-8146) is classified as A2 in the field of Biodiversity and A1 in Food Science.

Although Qualis' main objective is to evaluate the scientific publication of cross-country post-graduate programs, its use has been expanded to grants' concessions, promotions, admissions, among others. As the majority of publications from Brazilian researchers are in journals without citation metrics (Perlin et al. 2017), Qualis was created as a system to evaluate *all* journals with publications of Brazilian academics, including those without citation metrics. Several countries use similar evaluation systems, such as Australia, Colombia, Finland, Denmark, Germany, Pakistan, and India. As Qualis is an important driver of publications in Brazil, and not an isolated ranking system, an investigation regarding the reach of predatory journals is justified towards simultaneously better

understanding how they are related, and making our study more expandable to other countries. All the Qualis data used in this research was gathered from the CAPES website.⁸

Following the objective of analyzing the insertion and effect of Qualis in predatory practices, we estimate a control/treatment econometric model. The dependent variable is the number of publications for a particular journal, year and area of Qualis. The modelling approach was selected due to the counting nature of the dependent variable (Gujarati 2009). We use a generalized linear model with the *Poisson* distribution and Log link function as follows:

$$E(npub_{i,d,t}) = L(X_{i,d,t}) \tag{4}$$

$$X_{i,d,t} = \alpha_d + \beta_1 DPred_i + \beta_2 AftQ_{i,d,t} + \phi DPred_i * AftQ_{i,d,t} \tag{5}$$

where vector $npub_{i,d,t}$ is the number of publications for journal i in Qualis area d for year t . Variable $AfterQualis_{i,d,t}$ is another dummy that controls when journal i is inserted in Qualis of area d . Since the area of science from the Lattes database is not matchable with the area from Qualis, each publication is manually assigned a Qualis area based on a lookup table. This information is found by looking at each areas annual summary report available on the CAPES website. Since the production cycle of a scientific article takes time, we allow for a lag of one year in creating $AfterQualis_{i,d,t}$. This variable takes value one when journal i is in Qualis of area d of year $t - 2$. As with the previous model, we also acknowledge the heterogeneity of the number of publications of different scientific fields by estimating Eq. (4) with fixed effects for the Qualis area.

The interpretation of the model’s coefficients is straightforward. Coefficient β_1 measures the unconditional number of publications for predatory journals. We expect this parameter to be negative, as predatory publications are a small proportion of the dataset and are likely to have fewer publications than non-predatory journals. Coefficient β_2 quantifies the overall impact of a journal being added in Qualis, and we expect it to be positive. In the competition for academic funding and opportunities, Qualis sets the bar and it is a large factor in the decision of where to submit a paper. We thus expect that once a journal makes it in the Qualis of a particular field, it will be in greater demand from researchers from the same field.

The most important information from our econometric analysis is the sign and significance of ϕ . This parameter will show whether predatory journals receive more or fewer publications than non-predatory outlets once they are added to Qualis. Our expectation is that ϕ is positive, that is, once a predatory journal joins Qualis, it will receive more publications than others due to a lower acceptance standard. This is in line with the argument that researchers look for outlets with great rewards (Pyne 2017). If the journal is in Qualis, the scholar is less likely to challenge the quality and reach of the periodic. Moreover, the value of ϕ can show the potential of restricting the proliferation of predatory publications via filtering these cases in Qualis.

⁸ <http://www.capes.gov.br/avaliacao/qualis>.

Table 1 Descriptive statistics of publications

	Value
Number of publications	2,349,719
Number of researchers	102,969
Percentage of male researchers	53.84%
Percentage of researchers with PhD abroad	8.34%
Percentage of publications in Qualis	94.79%
Percentage of publications in SJR	34.87%
Percentage of publications in JCR	42.75%
Percentage of predatory (BEALL)	0.49%
Percentage of predatory (BEALL + DOAJ)	0.44%
Percentage of predatory (BEALL + DOAJ + IF)	0.26%

Results

Table 1 presents a statistical summary of our dataset. The gender of researchers is slightly biased towards males (53.84%). The proportion of researchers with a PhD obtained in an institution outside of Brazil is small, 8.34%. Unsurprisingly, a great proportion of the publications in the database are found in Qualis. As for international impact measures, we find that 34.87 and 42.75% of all publications are indexed in SJR and JCR, respectively.

Table 1 shows the proportion of predatory journals in our database for all years for each of the three classification schemes described in the earlier section. These percentages are relatively small, with 0.49% for the BEALL classification, 0.44% for BEALL + DOAJ and 0.26% for BEALL + DOAJ + IF. These small figures, however, can be deceiving, as predatory practices are a contemporary problem. In order to better understand the reach of predatory publishers over time, Fig. 1 shows the number and percentage of predatory publications over the years since 2000.

The number and proportion of predatory publications has increased significantly. The percentage for BEALL has tripled from 2010 to 2015. While the unconditional proportion

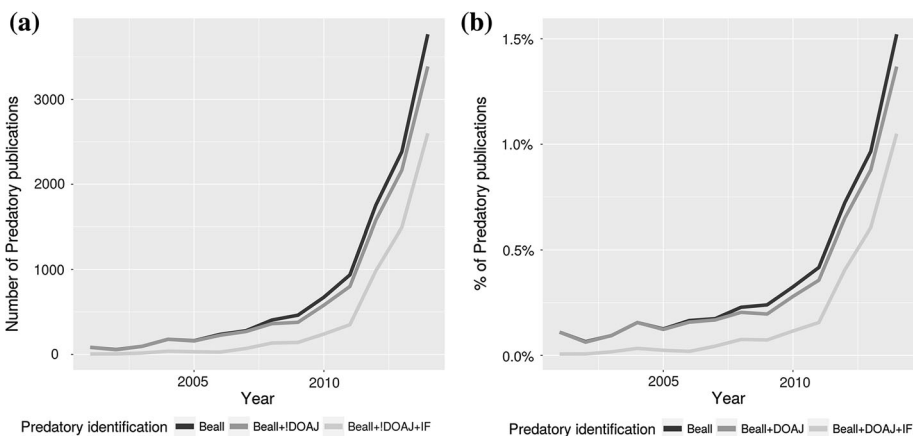


Fig. 1 The number and percentage of predatory publications per year. **a** Number of predatory publications, **b** percentage of predatory publications

of predatory publications is small, the tendency and exponential growth observed in the five past years is worthy of preoccupation.

Previous research has clearly shown that areas of science do research in different ways (Piro et al. 2013). Some, such as Medicine, have more options of international outlets and may place more weight in international publications. Other fields may lack the experience and opportunities to publish in established international publications. Unsurprisingly, we find that the tendency of increasing proportions of predatory publication differs when conditioning per field. Figure 2 presents the same information as Fig. 1, but restricting the relative proportions of predatory publications within each field of science.

While all areas show a recent rise in the proportion of predatory publications, Agricultural Sciences and Engineering stand out. It is difficult to specify the exact reasons, but we suspect that it might be due to the higher rewards granted within these areas for international publications (Pyne 2017). The proportions in Fig. 2 are also not that different when comparing the classifications schemes. The proportion of Beall + DOAJ + IF is almost the same as only Beall. These results clearly show that predatory publishing is becoming a systemic issue, and is not restricted to a particular area of science.

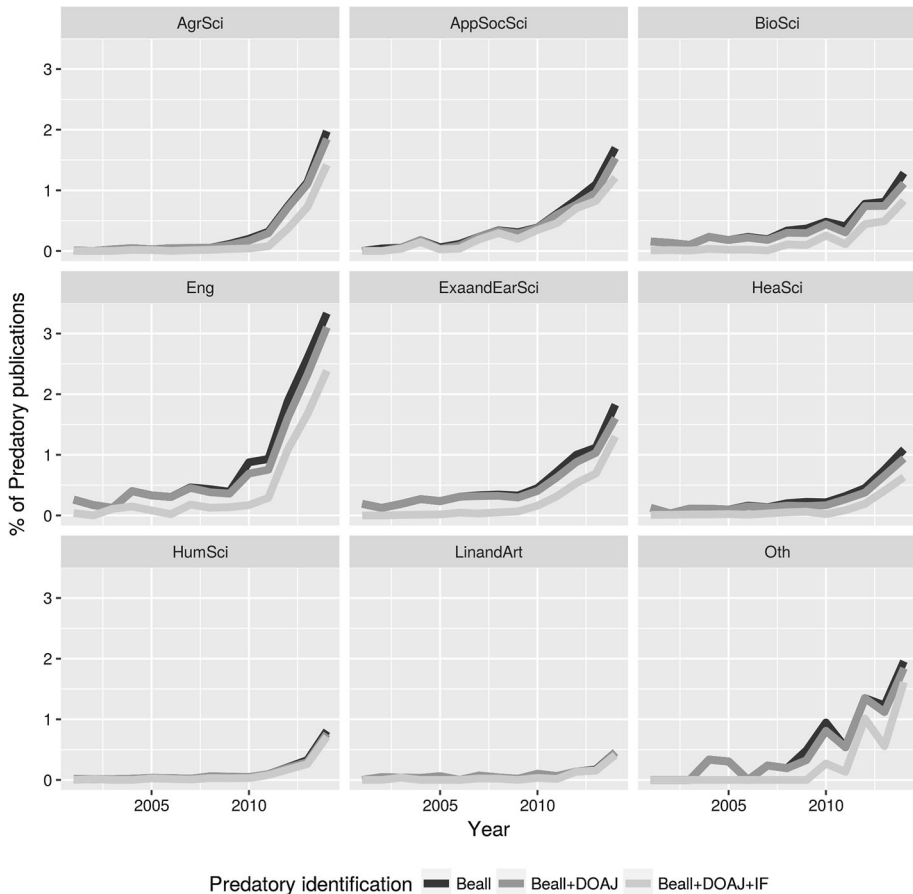


Fig. 2 Percentage of predatory publications by year and area

Figure 3 presents the percentage of predatory publications in different years of Qualis release. Following the results in Fig. 1, we see that the list of publishers and journals with questionable practices is increasingly being incorporated into Qualis, but in a linear fashion. We can further investigate the data by looking at each ranking of Qualis. Figure 4 shows the percentage of predatory journals per ranking in Qualis for all areas, across different years of Qualis release.

Figure 4 shows that most of the predatory journals are classified in the lower rankings, B4, B5 and C. Over the years, this proportion decreases as ranking increases, suggesting that Qualis assessment is having partial success in recognizing predatory venues. From 2013 onwards, we see a higher proportion of predatory journals registered in Qualis, possibly due to the proliferation of predatory publishers (Shen and Björk 2015). Nonetheless, we find predatory journals in all rankings of Qualis.

Econometric results

Our first model relates the probability of a publication in a predatory journal to a set of characteristics of authors. The estimated coefficients in Eq. (1) are presented in Table 2 for each analysis group. We find that researchers with PhDs obtained abroad and with publications in WoS are less likely to publish in predatory journals for all estimated models. The strength of the relationship, measured by the absolute value of β_1 , increases along the

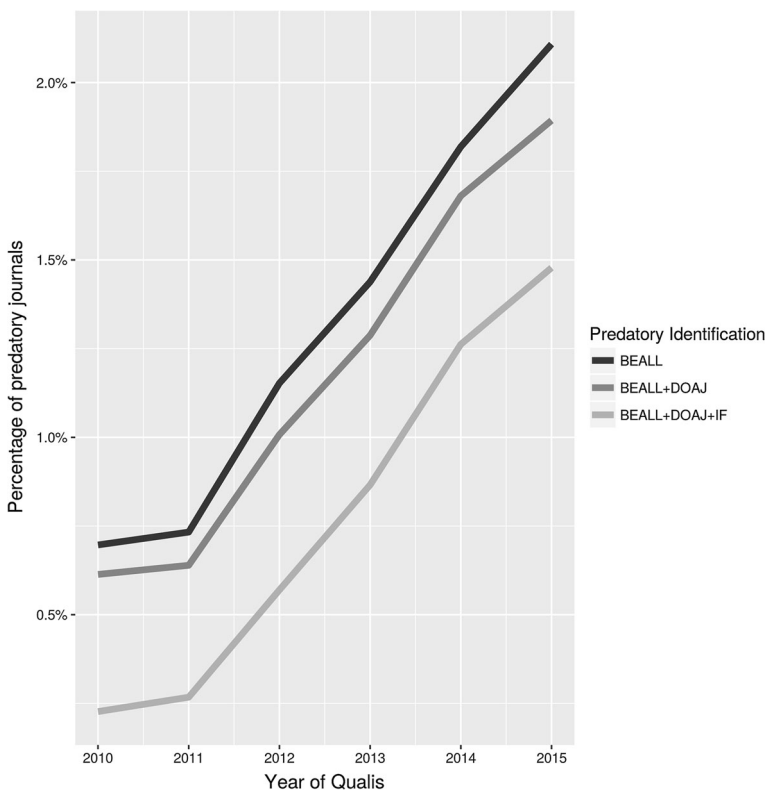


Fig. 3 Percentage of predatory publications in Qualis over the years

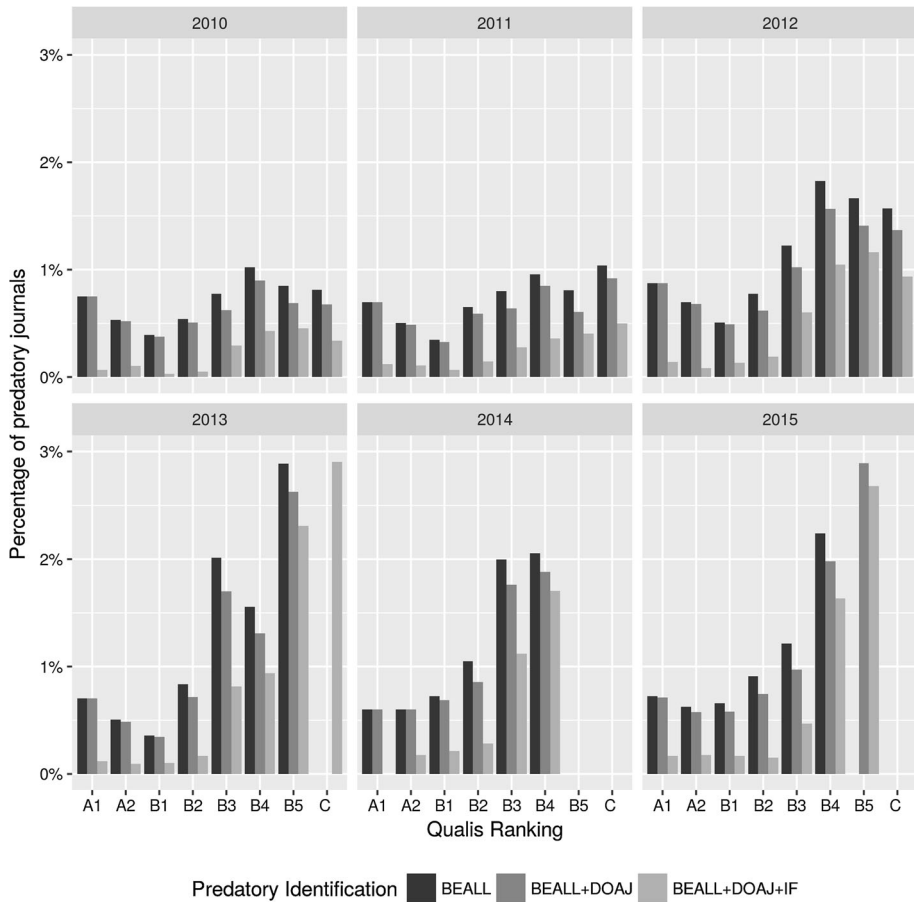


Fig. 4 Participation of predatory journals and publications in Qualis rankings

Table 2 Estimation results of Eq. 1

	BL	BL + DOA	BL + DOA+IF
DPhDAbroad	− 0.1645*** (0.0319)	− 0.1650*** (0.0334)	− 0.3372*** (0.0449)
DGender	0.0164 (0.0213)	0.0110 (0.0224)	− 0.0253 (0.0291)
Diff year	0.0156*** (0.0010)	0.0167*** (0.0011)	0.0226*** (0.0014)
N public (all)	0.0051*** (0.0003)	0.0053*** (0.0003)	0.0069*** (0.0003)
N public (WoS)	− 0.0079*** (0.0005)	− 0.0084*** (0.0005)	− 0.0121*** (0.0006)
LogLik	− 65054.2459	− 59455.3437	− 37815.7456
Num. obs.	2163822.0000	2163822.0000	2163822.0000

***, **, * Statistical significance at 1%, 5% and 10%

different classifications of predatory practices. This means that a scholar with a PhD obtained abroad is significantly less likely to publish in a journal that is not in Beall's list, DOAJ or the usual impact factor lists such as SJR and JCR. The lack of statistical significance for the gender parameter suggests that there is no evidence for gender bias towards predatory publications in the large publication dataset from Lattes.

As for the effect of career time, coefficient $DiffYear_i$, we report a positive relationship between the likelihood of a predatory article and the time interval from PhD completion to publication date. Likewise, the total number of publications ($NPub_i$) is also positively related to the variable of interest. This result strongly supports the lack of evidence for the argument that inexperienced researchers are driving predatory publications in Brazil. The data shows that experienced researchers, with significant career time and many publications in non-indexed journals, are publishing in predatory outlets. It is worth pointing out that such a result is consistent with the findings of Pyne (2017) and Wallace and Perri (2018), but based on a much larger publication sample.

As a robustness check, we also estimate model (1) only for the cases where the first author was taken into consideration. Certain academic collaboration schemes, where subsequent authors have not as significant direct participation or involvement in the work, could potentially bias our results. By restricting the data only for the first author, we are assuming that the first author is responsible for looking out for possible outlets and submitting papers on the journal website. As a consequence, it is very likely that the first author possesses more information about the journals, including the knowledge that the journal might have predatory editorial practices. However, the possibility of being unaware of this fact could nevertheless occur, with a higher probability for less-experienced academics (Xia et al. 2015; Al-Khatib 2016). The results from estimating model (1) in a subsample of first authors only are very similar to the results found in Table 2, with small changes in the values of the coefficients. All signs and statistical significance hold in this additional experiment.

Table 3 presents the estimation results of Eq. (5). Given our objective of finding a higher attractiveness of predatory journals once they enter Qualis, we restrict $npub_{i,d,t}$ only for journals that enter the higher rankings of Qualis (A1, A2, B1 and B2)⁹ First, we find parameter *Control* negative and significant at 1%. This result implies that, overall, predatory journals have fewer publications than non-predatory. We also find that coefficient *After* is significantly positive in all definitions of predatory publications. This implies that once a journal enters the Qualis of a particular area, it will have a significant amount of new publications from researchers within the same area. This result is not surprising, since Qualis indirectly defines academic rewards and the allocation of public funds.

The most important result of Table 3 is for the interaction of *Control* and *After*. Coefficient ϕ tells us whether a predatory journal publishes more papers than other non-predatory ones once it is included in Lattes. We find that parameter ϕ is positive as expected for all cases, with statistical significance at 1%. Interestingly, the value of ϕ in BL + DOAJ+IF is more than twice the value found for BL and BL + DOAJ. This result clearly shows how journals that are more likely to have predatory practices are being favored by authors when they enter Qualis, which explains the significant growth of these journals over time.

⁹ The results for all rankings are very comparable, with matching coefficient signs and significance. However, the absolute value decreases significantly.

Table 3 Estimation results of Qualis model, Eq. (4)

	BL	BL + DOAJ	BL + DOAJ+IF
Control	– 0.8450*** (0.0244)	– 0.8446*** (0.0250)	– 1.3921*** (0.0796)
After	0.2913*** (0.0020)	0.2913*** (0.0020)	0.2928*** (0.0020)
Control * after	0.7918*** (0.0368)	0.8242*** (0.0373)	1.7698*** (0.0939)
LogLik	– 1531917.6408	– 1531956.1826	– 1532443.1892
Num. obs.	159283.0000	159283.0000	159283.0000

***, **, * Statistical significance at 1%, 5% and 10%

Discussion and policy implications

In general, the obtained results followed the major findings from previous studies in predatory publishing. Predatory publications, under three different classification groups, are still a small proportion of all publications of Brazilian researchers from 2000 to 2015. However, the rate of growth of publications in journals with questionable practices is higher than the same rate for publications considering all journals, resulting in an unsettling exponential increase in the production of predatory publications.

One of the key factor that could help explain the exponential growth rate of publications in predatory outlets is the substantial increase in the local incentives for international publications by governmental agencies. Brazil has significantly ramped up its expenditure on research in recent years. Brazil has public investments in research, around 0.61% GDP, very close to the average of OECD countries (0.69%). Nowadays, it is very difficult to obtain a research grant in the Brazilian academy without regular publication in journals categorized by Qualis/Capes in rankings A1 or A2. As several journals on Beall’s list are evaluated in these two categories, the financial rewards of predatory publishing become very interesting in comparison with prestigious journals for each discipline.

Unlike the case analyzed by Omobowale et al. (2014) in Nigeria, we find no evidence that predatory publications replaces local publications. In our study, the most part of publications were not listed in BL indeed. As Brazil has a longer tradition in publications, the replacement is not likely to happen as the scientific institutions are not similar. However, the fast growing of predatory publications found in our dataset indicates that there might have some gain achieved through the publishing in this kind of outlet (BL, in our case), which are consistent with previous studies (Pyne 2017). Hiring, job promotions, financial compensations, funding, pressures to publish and internationalization of publications (by non-native english speakers) are some of the possible explanations to the phenomena in the literature. In Brazil, as in other non-native english developing countries, publishing in international outlets might have incentives.

By analyzing the profile of researchers that are likely to publish in predatory journals, we find a significant relationship between an international experience and likelihood of publishing in predatory journals. Those with a PhD obtained abroad and with more publications in indexed journals are less likely to publish in journals with questionable predatory practices. Once again, we show the importance of funding scholars to study abroad as pointed out in a previous study (Perlin et al. 2017). The possibility of studying in

mature academic institutes seems to have a good effect on the international reach of the publication of Brazilian academics. This result suggests another benefit of increasing the public funding of PhD scholarships abroad, at least until all research works in all disciplines can be considered to be of the same level as those carried out in countries with a solid academic history and reputation.

Although our research indicates that PhD obtained abroad are less likely to publish in journals with questionable predatory practices, Wallace and Perri (2018) results shows that this point should be considered with precautions. As experienced researchers in developed countries also publishes in some probable predatory outlets listed in BL, the role of academic formation is only a partial element to explain the profile. Employment pressures and the high competition to publish in traditional outlets might force or encourage experienced academics to publish in controversial outlets or to misconduct. However, work conditions is a difficult type of factor to obtain in our dataset.

Surprisingly, in contrast to prior studies such as Frandsen (2017), Al-Khatib (2016) and Xia et al. (2015), our study shows that young and inexperienced researchers are not necessarily the profile of those who publishes in predatory journals. This raises new questions regarding who chooses to publish in these venues. The potential of academic misconduct in these cases can be more likely to happen as researchers are aware of predatory practices and might take gains from it, which is consistent with Pyne (2017). Thus, the academic context make experienced and non-experienced alike to publish in predatory outlets. This result is also consistent with Wallace and Perri (2018) in the field of Economics.

When looking at Qualis, we find that predatory journals take part in all rankings. Further econometric analysis also shows that predatory journals, when they enter Qualis, are significantly more in-demand than other journals, evidencing the better financial rewards of predatory publishing in comparison with non-predatory ones previously identified by Pyne (2017) in a small case study. In the existence of official academic ranking system, Qualis in the case of our study, the inclusion of predatory journals will create incentives to publish in this kind of outlet. As predatory practices are reported to be uncritical in the reviewing process, this generates concerns regarding possible incentives for scientific misconduct too.

As publications in questionable journals drain a significant amount of mostly public research resources, there is an urgent necessity of better regulating this issue. As pointed out by Omobowale et al. (2014), the predatory publishers inform the global view of the perception of their authors. Excessive publication in predatory publishing characterizes the local research as sub-standard and weak in the local academy, wasting significant public resources in a country that faces serious education, health, and violence problems. Therefore, decreasing the importance of predatory publications in the academic career is becoming an urgent necessity. The first step in battling predatory journals is in identifying them. With that in mind, we provide a list of ISSNs and names of journals that are found by our software. The data is provided as a .csv file that can be used to check whether a particular ISSN/journal is predatory or not.

Given the *leverage* effect of Qualis in the number of publications in predatory journals, we strongly suggest that the compiled list be used in identifying allegedly predatory journals in the categorization of Qualis. The provided data should be used as a flag for further, qualitative, inspection. A large-scale check process based on names and ISSNs available within the .csv file can be easily carried out. We suggest that, after a positive identification, the merits of the journal should be analyzed following [DOAJ principles](#).

We believe that Qualis could have more transparent procedures regarding the definition of what is considered an international, scientific, and peer-reviewed journal. We must

recognize that no financial and publication model is perfect, and cannot be morally superior to others or capable of assuring high-quality (Haug 2013), since all models are vulnerable to exploitation and misbehavior. However, with greater transparency, based on well-defined criteria and experienced researchers, it is possible to make the academic community understand that there are different types of publication, and that not all journals can be referred to as “scientific”. Doing so sends a strong message to the community that journals with such practices are not going to be considered for research evaluation and should be avoided if the researcher has such objectives. We believe that a strong action from CAPES, regarding Qualis, is the most cost-effective and direct way of dealing the proliferation of predatory publishers and journals.

It should be noted that we are not claiming that these journals should be banished, since they can be alternative options for publication of a massive number of new authors or a place for discussion among peers. After all, the relevance of a publication is given by its readers. We are simply claiming that researchers should be made aware that it can be a waste of resources to publish in such journals from a purely scientific point of view.

The empirical findings of this study are restricted to the Brazilian academic system. However, its results and tools can be broadly applied, mainly for countries with similar academic levels. The use of a local journal ranking system, such as Qualis, is not new. Other countries such as Australia with the ERA (Excellence in Research) system, Norway with the CRISTin initiative, and Colombia with Publindex have their own standards towards local quality assessments. Given the international reach of predatory publishers, the penetration of predatory journals is not likely to be solely restricted to Qualis. We suspect that empirical work that relates predatory journals and local ranking systems in other countries, with the tools provided by this work, can obtain interesting highlights about predatory publishing worldwide.

Conclusions

In this paper, we look into the extent of predatory publications in a large database of scientific publications from Lattes. We identify predatory outlets using Beall’s list, the DOAJ (directory of open access journals), and standard impact factors, SJR and JCR. We find that allegedly predatory journals are a small but exponentially-increasing part of all publications. Such a result is essential for the identification methods of predatory journals and different fields of science.

Our econometric analysis shows that the expected profile of academics who are likely to publish in predatory journals are those with doctorates obtained in local institutions and with significant career time. The bibliometric data does not support the idea that new and inexperienced scholars are the only ones publishing in these venues. We also find that predatory journals are part of the local journal ranking system, Qualis. Based on an econometric model, we test the hypothesis that predatory journals are prioritized by authors when they enter Qualis. We find a positive result, meaning that predatory journals show a higher increase of publications when they are inserted in Qualis.

The main contribution of this research is to introduce a quantitative study based on a large scale database, producing a comprehensive analysis of the profile of researchers publishing in potential, possible, or probable predatory scholarly journals. To the best of our knowledge, this is the first paper to simultaneously cover all scientific areas, using a massive database of publishing papers, and to apply an econometric model to analyze data.

We strongly believe that our study can bring important highlights about predatory publishing, a problem with diverse multinational and multicultural sources.

Although our study was focused on Brazil, the quantitative approach can be easily implemented in other developing countries, with a similar academia structure and development stage. In this sense, the results of this study might raise new questions about the international reach of predatory publishers in other countries, following the Brazilian case. The most worrying result is the backup of official governmental scientific production classification to allegedly predatory publications, as the case of Qualis. We advocate that government research agencies use a transparent method for defining international, scientific, peer-reviewed papers, making researchers choose the objective of their publication and avoiding predatory journals as scientific outlets. This paper provides tools that can help government agencies and scholars to identify predatory publications and the profile of researchers which are using this outlet to disseminate their research. We hope that the large-scale use of these tools can help decrease the harm that predatory publishers are causing in the academic system, and in the OA publishing model.

References

- Abbasi, K. (2012). The debate around open-access publishing. *Journal of the Royal Society of Medicine*, 105, 185.
- Abramo, G., Cicero, T., & D'Angelo, C. A. (2015). Should the research performance of scientists be distinguished by gender? *Journal of Informetrics*, 9(1), 25–38.
- Al-Khatib, A. (2016). Protecting authors from predatory journals and publishers. *Publishing Research Quarterly*, 32(4), 281–285.
- Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, 489(7415), 179.
- Beall, J. (2013). Medical publishing triage-chronicling predatory open access publishers. *Annals of Medicine and Surgery*, 2(2), 47–49.
- Beall, J. (2015a). *Criteria for determining predatory open-access publishers*. http://publica.upc.edu/sites/default/files/arxius_site/criteris_editorials_fraudulent.pdf. Accessed 5 May 2017.
- Beall, J. (2015b). Predatory journals and the breakdown of research cultures. *Information Development*, 31(5), 473–476.
- Beall, J. (2017). What i learned from predatory publishers. *Biochemia Medica: Biochemia Medica*, 27(2), 273–278.
- Berger, M., & Cirasella, J. (2015). Beyond Beall's list better understanding predatory publishers. *College and Research Libraries News*, 76(3), 132–135.
- Bowman, J. D. (2014). Predatory publishing, questionable peer review, and fraudulent conferences. *American Journal of Pharmaceutical Education*, 78(10), Article 176.
- Brooks, C., Fenton, E. M., & Walker, J. T. (2014). Gender and the evaluation of research. *Research Policy*, 43(6), 990–1001.
- Brostrm, G. (2017). *glmmML: Generalized linear models with clustering*. <https://CRAN.R-project.org/package=glmmML>, r package version 1.0.2.
- Butler, D. (2013). The dark side of publishing. *Nature*, 495(7442), 433.
- Clark, A. M., & Thompson, D. R. (2016). Five (bad) reasons to publish your research in predatory journals. *Journal of Advanced Nursing*. <https://doi.org/10.1111/jan.13090>.
- Djuric, D. (2015). Penetrating the omerta of predatory publishing: The Romanian connection. *Science and Engineering Ethics*, 21(1), 183–202.
- Frandsen, T. F. (2017). Are predatory journals undermining the credibility of science? A bibliometric analysis of citers. *Scientometrics*. <https://doi.org/10.1007/s11192-017-2520-x>.
- Gujarati, D. N. (2009). *Basic econometrics*. New York: Tata McGraw-Hill Education.
- Haug, C. (2013). The downside of open-access publishing. *New England Journal of Medicine*, 368(9), 791–793.
- Kulczycki, E. (2017). The rise of predatory journals: The case of Dr. Anna fraud. *NAUKA*, 1(3), 71–83.
- Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B. C., & Hedlund, T. (2011). The development of open access journal publishing from 1993 to 2009. *PLoS ONE*, 6(6), e20,961.

- Lukić, T., Blešić, I., Basarin, B., Ivanović, B. L., Milošević, D., & Sakulski, D. (2014). Predatory and fake scientific journals/publishers: A global outbreak with rising trend: A review. *Geographica Pannonica*, 18(3), 69–81.
- Manca, A., Martínez, G., Cugusi, L., Dragone, D., Dvir, Z., & Deriu, F. (2017a). The surge of predatory open-access in neurosciences and neurology. *Neuroscience*, 353, 166–173.
- Manca, A., Martínez, G., Cugusi, L., Dragone, D., Mercuro, G., & Deriu, F. (2017b). Predatory open access in rehabilitation. *Archives of Physical Medicine and Rehabilitation*, 98(5), 1051–1056.
- Nwagwu, W. E. (2016). Open access in the developing regions: Situating the altercations about predatory publishing/l'accès libre dans les régions en voie de développement: Situation de la controverse concernant les pratiques d'édition déloyales. *Canadian Journal of Information and Library Science*, 40(1), 58–80.
- Omobowale, A. O., Akanle, O., Adeniran, A. I., & Adegboyega, K. (2014). Peripheral scholarship and the context of foreign paid publishing in nigeria. *Current Sociology*, 62(5), 666–684.
- Perlin, M. (2017). *Predatory: Tools for detecting predatory publishers and journals*. <https://CRAN.R-project.org/package=predatory>, r package version 1.2.
- Perlin, M. S., Santos, A. A., Imasato, T., Borenstein, D., & Da Silva, S. (2017). The Brazilian scientific output published in journals: A study based on a large cv database. *Journal of Informetrics*, 11(1), 18–31.
- Piro, F. N., Aksnes, D. W., & Rrstad, K. (2013). A macro analysis of productivity differences across fields: Challenges in the measurement of scientific publishing. *JASIST*, 64, 307–320.
- Pyne, D. (2017). The rewards of predatory publications at a small business school. *Journal of Scholarly Publishing*, 48(3), 137–160.
- Rørstad, K., & Aksnes, D. W. (2015). Publication rate expressed by age, gender and academic position—a large-scale analysis of Norwegian academic staff. *Journal of Informetrics*, 9(2), 317–333.
- Shen, C., & Björk, B. C. (2015). 'Predatory' open access: A longitudinal study of article volumes and market characteristics. *BMC Medicine*, 13(1), 1.
- Shuva, N. Z., & Taisir, R. (2016). Faculty members perceptions and use of open access journals: Bangladesh perspective. *IFLA Journal*, 42(1), 36–48.
- Sorokowski, P., Kulczycki, E., Sorokowska, A., & Pisanski, K. (2017). Predatory journals recruit fake editor. *Nature*, 543, 481–483.
- Strielkowski, W. (2017). Predatory publishing: What are the alternatives to beall's list? *The American Journal of Medicine*. <http://www.sciencedirect.com/science/article/pii/S0002934317311968>.
- Van Noorden, R. (2013). Brazilian citation scheme ousted. *Nature*, 500(7464), 510–1.
- Wallace, F. H., & Perri, T. J. (2018). Economists behaving badly: Publications in predatory journals. *Scientometrics*. <https://doi.org/10.1007/s11192-018-2690-1>.
- Xia, J. (2015). Predatory journals and their article publishing charges. *Learned Publishing*, 28(1), 69–74.
- Xia, J., Harmon, J. L., Connolly, K. G., Donnelly, R. M., Anderson, M. R., & Howard, H. A. (2015). Who publishes in “predatory” journals? *Journal of the Association for Information Science and Technology*, 66(7), 1406–1417. <https://doi.org/10.1002/asi.23265>.